

BLIND COMPONENT SEPARATION IN WAVELET SPACE. APPLICATION TO CMB ANALYSIS.

Y. Moudden¹, J.-F. Cardoso², J.-L. Starck¹ and J. Delabrouille³

¹ DAPNIA/SEDI-SAP, CEA/Saclay, F-91191 Gif-sur-Yvette, France

² CNRS/ENST, 46 rue Barrault, F-75634 Paris, France

³ CNRS/PCC, Collège de France, 11 place Marcelin Berthelot, F-75231 Paris, France

yassir.moudden@cea.fr, cardoso@tsi.enst.fr, jstarck@cea.fr, delabrouille@cdf.in2p3.fr

ABSTRACT

It is a recurrent issue in astronomical data analysis that observations are unevenly sampled or incomplete maps with missing patches or intentionally masked parts. In addition, many astrophysical emissions are non stationary processes over the sky. Hence spectral estimation using standard Fourier transforms is no longer reliable. Spectral matching ICA (SMICA) is a source separation method based on covariance matching in Fourier space which is successfully used for the separation of diffuse astrophysical emissions in Cosmic Microwave Background observations. We show here that wavelets, which are standard tools in processing non stationary data, can profitably be used to extend SMICA. Among possible applications, it is shown that gaps in data are dealt with more conveniently and with better results using this extension, wSMICA, in place of the original SMICA. The performances of these two methods are compared on simulated CMB data sets, demonstrating the advantageous use of wavelets.

Keywords : *blind source separation, cosmic microwave background, wavelets, data analysis*

1. INTRODUCTION

The detection of Cosmic Microwave Background (CMB) anisotropies on the sky has been over the past three decades subject of intense activity in the cosmology community.

The CMB, discovered in 1965 by Penzias and Wilson, is a relic radiation emitted some 13 billion years ago, when the Universe was about 370.000 years old. Small fluctuations of this emission, tracing the seeds of the primordial homogeneities which gave rise to present large scale structures as galaxies and clusters of galaxies, have been observed by a number of experiments such as Archeops [16], Boomerang [17], Maxima [18] and WMAP [19].

The precise measurement of these fluctuations is of utmost importance for Cosmology. Their statistical properties (spatial power spectrum, Gaussianity) strongly depend upon the cosmological scenarios describing the properties and evolution of our Universe as a whole, and thus permit to constrain these models as well as to measure the cosmological parameters describing the matter content, the geometry, and the evolution of our Universe [20].

Accessing this information, however, requires disentangling in the data the contribution of several distinct astrophysical sources, all of which emit radiation in the frequency range used for CMB observations [21]. This problem of component separation, in the field of CMB studies, has thus been the object of many dedicated studies in the past.

To first order, the total sky emission is modelled as a linear mixture of a few independent processes. The observation of the sky with detector d is then a noisy linear mixture of

N_c components :

$$y_d(\theta, \phi) = \sum_{j=1}^{N_c} A_{dj} s_j(\theta, \phi) + n_d(\theta, \phi) \quad (1)$$

where s_j is the emission template for the j th astrophysical process, herein referred to as a *source* or a *component*. The coefficients A_{dj} reflect emission laws while n_d accounts for noise. When N_d detectors provide independent observations, this equation can be put in vector-matrix form :

$$X(\theta, \phi) = AS(\theta, \phi) + N(\theta, \phi) \quad (2)$$

where X and N are vectors of length N_d , S is a vector of length N_c , and A is the $N_d \times N_c$ mixing matrix.

Given the observations of such a set of independent detectors, component separation consists in recovering estimates of the maps of the sources $s_j(\theta, \phi)$. Explicit component separation has been investigated first in CMB applications by [22], [21], and [23]. In these applications, recovering component maps is the primary target, and all the parameters of the model (mixing matrix A_{dj} , noise levels, statistics of the components, including the spatial power spectra) are assumed to be known and used as priors to invert the linear system.

Recent research has addressed the case of an imperfectly known mixing matrix. It is then necessary, to estimate it (or at least some of its entries) directly from the data. For instance, Tegmark *et al.* assume power law emission spectra for all components except CMB and SZ, and fit spectral indices to the observations [15]. More recently, blind source separation or independent component analysis (ICA) methods have been implemented specifically for CMB studies. The work of [2], further extended by [4] implements a blind source separation method exploiting the non-Gaussianity of the sources for their separation, which permits to recover the mixing matrix A and the maps of the sources.

Delabrouille *et al.* [1] propose an approach exploiting the spectral diversity of components, with the new point of view that spatial power spectra are actually the main unknown parameters of interest for CMB observations. The estimation of a set of parameters of the model, among which the spatial power spectra of the components, is made using a set of band-averaged spectral covariance matrices in Fourier space.

While working in the Fourier domain has a number of advantages, it also has a number of drawbacks. When components or noise are strongly non-stationary, one may wish to avoid the averaging induced by Fourier transforms. In addition, when dealing with real-life observations, quite often the coverage is incomplete for a reason or another. Either the instrument observes only a fraction of the sky, or some regions of the sky have to be rejected due to localised strong astrophysical sources of contamination : compact radiosources or galaxies, strong emitting regions in the galactic plane.

Blind component separation (and in particular estimation of the mixing matrix), as discussed by Cardoso [9], can be achieved in several different ways. The first of these exploits non-Gaussianity of all but possibly one components. However, this is not recommended for mixtures where one component is close to Gaussian and all observations suffer from additive Gaussian noise. The component separation method of Baccigalupi [2] and Maino [4] is based on this method. The second, which exploits spectral diversity (or non-stationarity in Fourier domain), has the advantage that detector-dependent beams can be handled easily, since the convolution with a point spread function in direct space becomes a simple product in Fourier space. SMICA is an extension of this approach to noisy observations. Finally, the third method exploits non-stationarity in real space. It is adapted to situations where components are strongly non-stationary in real space.

As an extension of these last two methods, it is natural to investigate the possible benefits of exploiting both non-stationarity and spectral diversity for blind component separation using wavelets. Indeed wavelets are powerful tools in revealing the spectral content of non-stationary data. In what follows, we first recall in section 2 the fundamental principles of Spectral Matching ICA. Then, after a brief reminder of the *à trous* wavelet transform, we discuss in section 3 the extension of SMICA for component separation in wavelet space in order to deal with non-stationary data. Considering the problem of incomplete data as a model case of practical significance for the comparison of SMICA and its extension wSMICA, numerical experiments and results are reported in section 4. From these, conclusions are drawn in section 5.

2. SMICA

This paragraph recalls the main hypotheses and equations of the SMICA algorithm which we actually extended to deal with gapped data. For ease of presentation, we concentrate on the 1D case since the extension to two dimensional data is straightforward. Detailed descriptions and discussions of this method can be found in [8, 11] and results of previous applications to CMB analysis can be read in [1, 7].

2.1 Model and cost function

Spectral matching ICA is a blind source separation technique that overcomes the inseparability of Gaussian sources using standard ICA methods by relying on their assumed spectral diversity : SMICA allows us to recover independent Gaussian colored sources from observed noisy mixtures provided their spectra are substantially not proportional [14].

Considering the linear instantaneous mixing model with additive noise defined by (2), with the assumption that noise and source processes are centered, stationary and independent, and denoting $R_X(\nu)$, $R_S(\nu)$ and $R_N(\nu)$ the spectral covariances of X , S and N respectively, it follows from (2) that for any value of the reduced frequency $\nu \in [-0.5, 0.5]$,

$$R_X(\nu) = AR_S(\nu)A^\dagger + R_N(\nu) \quad (3)$$

when we further assume independence between source and noise processes. Clearly, independence also implies that $R_S(\nu)$ and $R_N(\nu)$ are diagonal matrices.

Given a batch of T regularly spaced experimental data samples $X_{t=1 \rightarrow T}$ and a set $\{\nu_{q,q=1 \rightarrow Q}\}$ of Q different reduced frequencies chosen *a priori*, estimates $\hat{R}_X(\nu_q)$ of $R_X(\nu_q)$ of the spectral covariance at these frequencies can be computed easily in a number of ways. The basic idea of spectral matching is to fit the model covariances of equation (3) to these

experimental covariances by minimizing, over all or a subset of the model parameters $\theta = \{R_S(\nu_q), R_N(\nu_q), A\}$, the functional

$$\phi(\theta) = \sum_{q=1}^Q \alpha_q \mathcal{D}(\hat{R}_X(\nu_q), AR_S(\nu_q)A^\dagger + R_N(\nu_q)) \quad (4)$$

where $\mathcal{D}(\cdot, \cdot)$ is a measure of the divergence between two covariance matrices, and α_q are weights which depend on q . This adjustment results in estimates $\hat{\theta} = \{\hat{R}_S(\nu_q), \hat{R}_N(\nu_q), \hat{A}\}$ of the model parameters and hence enables us to achieve the desired source separation. It is worth highlighting that resorting to covariances highly reduces data dimension, which is of great interest to astrophysical applications where data sets tend to become very large. Moreover, it may be argued in the stationary Gaussian case that this reduction is without significant loss of information [1].

Although any reasonable set of weights α_q and divergence $\mathcal{D}(\cdot, \cdot)$ can be used in (4) to assess spectral mismatch, this will affect the statistical properties of the estimated model parameters $\hat{\theta} = \{\hat{R}_S(\nu_q), \hat{R}_N(\nu_q), \hat{A}\}$. Deriving a mismatch criterion from higher statistical principles such as maximum likelihood should lead to better such estimates.

In the SMICA method, the divergence \mathcal{D} used is given by

$$\mathcal{D}_{KL}(R_1, R_2) = \frac{1}{2} \left(\text{Tr}(R_1 R_2^{-1}) - \log \det(R_1 R_2^{-1}) - m \right) \quad (5)$$

which actually derives from the Kullback-Leibler divergence between two centered Gaussian distributions with size $m \times m$ covariance matrices R_1 and R_2 . Moreover, assuming constant source $R_{S,q}^f$ and noise $R_{N,q}^f$ power spectra, over frequency domains $\{F_q\}_{q \in [1, Q]}$, SMICA uses refined unbiased estimates $\hat{R}_{X,q}^f$ of the mixture covariance matrices $R_{X,q}$ defined by :

$$\hat{R}_{X,q}^f = \frac{1}{n_q} \sum_{p=0, \frac{p}{T} \in F_q}^{T-1} \tilde{X}(\frac{p}{T}) \tilde{X}(\frac{p}{T})^\dagger \quad (6)$$

where \tilde{X} is the discrete Fourier transform of X ,

$$\tilde{X}(\nu) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} X(t) e^{-2\pi j \nu t}, \quad (7)$$

the F_q are non-overlapping domains in $[-1/2, 1/2]$, symmetric with respect to zero, with their positive parts centered on ν_q , and n_q is the number of $\frac{p}{T}$ that fall in F_q . It follows from this definition that the entries of $\hat{R}_{X,q}^f$ are in fact all real. The statistical grounds and implications of these choices are explored in [8, 14] where it is shown that SMICA can be derived asymptotically from the maximum likelihood principle in the particular case of stationary processes in the Whittle approximation. This latter approximation asserts that the Fourier coefficients $\tilde{X}(\frac{p}{T})$ of a stationary process $X(t)$ are asymptotically Gaussian, uncorrelated, centered with spectral covariance equal to $R_X(\frac{p}{T})$.

As a result, the model covariance (3) is finally rewritten as :

$$R_{X,q}^f = AR_{S,q}^f A^\dagger + R_{N,q}^f \quad (8)$$

and the derived spectral matching criterion is given by

$$\phi(\theta) = \sum_{q=1}^Q n_q \mathcal{D}_{KL}(\hat{R}_{X,q}^f, AR_{S,q}^f A^\dagger + R_{N,q}^f) \quad (9)$$

to be minimized with respect to the new set of parameters $\theta = (A, R_{S,q}^f, R_{N,q}^f)$.

The previous definitions are easily extended for the method to be applied to real images. The above F_q are naturally replaced by 2D domains in the frequency plane [11]. These are best chosen, based on available prior information relative to source spectra, to enhance spectral diversity [14]. Regarding our application to CMB analysis, the supposed spatial stationarity and isotropy of the sources strongly suggests taking rings centered on the null frequency which are finally simply described as 1D frequency bands.

An especially important limiting case, for simulation purposes, is when the mixing matrix is square and invertible, and when the mixtures are assumed without noise. Then, as shown in [8], the likelihood can be directly related to a joint diagonalization criterion of spectral covariance matrices for which an efficient optimization algorithm is actually available.

2.2 Parameter optimization

Finding the model closest to the data in the sense of SMICA's objective function benefits from the latter's connection to the maximum likelihood principle and indeed the EM algorithm is shown to be a fruitful search method in [1] where it is fully described. Actually, this latter algorithm was slightly modified in order to deal with the case of colored noise N in (2). Another useful enhancement was to allow for constraints to be set on the model parameters so that prior information such as bounds on some entries of the mixing matrix A could be included. The details of this constrained EM algorithm are given in appendix A.

Eventually, using the EM algorithm in simulation, it appeared that after a quick start, convergence slowed down dramatically in a second stage possibly owing to poor signal to noise ratio in some frequency bands. In order to speed convergence back up, it was found profitable to alternately use fixed numbers of EM steps and BFGS steps [1, 7] in a heuristic procedure.

An unavoidable issue in optimization is that of initiating the search method and this, obviously, is most critical when the function to be optimized is strongly suspected to be multimodal. Such may very well be the case with (4). This point though is left aside in what follows since our prime interest is in the study of the statistical performances of different estimators of the model parameters θ . In the simulations discussed further down, the optimal values of the parameters are sought starting from the true mixing matrix and the spectral covariances estimated from the initial separate source and noise maps.

2.3 Component map estimation

As by-products of the SMICA method, estimates $\hat{R}_{S,q}^f$ and $\hat{R}_{N,q}^f$ of the different signal and noise covariances are obtained in the model fitting step and can be used for reconstructing the source maps *via* Wiener filtering the data maps in Fourier space, in each frequency band $\nu \in F_q$:

$$\hat{S}(\nu) = (\hat{A}^\dagger \hat{R}_{N,q}^{f-1} \hat{A} + \hat{R}_{S,q}^{f-1})^{-1} \hat{A}^\dagger \hat{R}_{N,q}^{f-1} X(\nu) \quad (10)$$

In the limiting case where noise is small compared to signal components, $\hat{R}_{S,q}^{f-1}$ is negligible and the above filter reduces to

$$\hat{S}(\nu) = (\hat{A}^\dagger \hat{R}_{N,q}^{f-1} \hat{A})^{-1} \hat{A}^\dagger \hat{R}_{N,q}^{f-1} X(\nu) \quad (11)$$

which is also the generalized least square solution under Gaussian statistics.

Note however that the Wiener filter is only one possibility among others for inverting (2). Its optimality is true in the restricted case of Gaussian noise and signal processes. In real case applications, other inverting schemes should also be experimented [1].

3. WAVELETS AND SMICA

The SMICA method for spectral matching in Fourier space has proven to be a very powerful tool for CMB spectral estimation in multidetector experiments. It is particularly useful to identify and remove residuals of poorly known correlated systematics and astrophysical foreground emissions contaminating CMB maps. However, SMICA suffers from several practical difficulties when dealing with real data.

Indeed, actual components are known to depart slightly from the ideal linear mixture of equation (2). The mixing matrix (in particular those columns of A which correspond to galactic emissions) is known to depend somewhat on the direction of observation or on spatial frequency. Measuring the dependence $A(\theta, \phi)$ is of interest for future experiments as Planck, and can not be achieved directly with SMICA. Further, the components are known to be both correlated and non stationary. For instance, galactic dust emissions are strongly peaked towards the galactic plane. A Fourier (or spherical harmonics) transform inevitably mixes contributions from high galactic sky, nearly free of foreground contamination, and contributions from within the galactic plane. Noise levels themselves may be quite non stationary, with high SNR regions observed for a long time and low SNR regions poorly observed.

When there are sharp edges on the maps or gaps in the data, corresponding to unobserved or masked regions, spectral estimation using the periodogram or the Daniell-like smoothed periodogram as in (6) is also not the most satisfactory procedure. Although apodizing windows may help cope with edge effects in Fourier analysis, they are not very straightforward to use in the case of arbitrarily shaped 2D maps with arbitrarily shaped 2D gaps, such as provided by the Archeops experiment [16]. Clearly, the spectral analysis of gapped data requires tools different from those used to process full data sets, if only because the hypothesized stationarity of the data is greatly disturbed by the missing samples.

Common such methods often amount to first trying to fill the gaps with estimates of the missing samples and then using standard spectral estimators. However, the data interpolation stage is critical and cannot be completed without prior assumptions on the data [12]. We preferred to rely on methods intrinsically dedicated to the analysis of non-stationary data such as the wavelet transform, widely used to reveal variations in the spectral content of time series or images, as they permit to single out regions in direct space while retaining localization in the frequency domain. We see next how to reformulate (4) so to take advantage of wavelet transforms when dealing with non-stationary data. A particular case in which wavelets are shown to be an especially powerful tool is that of incomplete data. Note that in what follows, the locations of the missing samples are always known.

3.1 Wavelet transform : the *à trous* algorithm

We give here the necessary background on the *à trous* algorithm which, among the several possible wavelet transform implementations, is the one we retained in our simulations. With the compact supported cubic B_3 spline

as scaling function $\phi(k)$, or its 2D quasi-isotropic extension $\phi(k)\phi(l)$, the *à trous* algorithm has been shown to be well suited to the analysis of atrophysical data where translation invariance is desirable and the accent is seldom set on data compression [10]. For this choice of scaling function, the scaling equation (13) is satisfied and therefore fast implementations of the decomposition and reconstruction steps of the *à trous* transform are available [10].

Consider for instance a sampled 1D signal $c_0(k)$ of length T . The *à trous* algorithm recursively produces smoother approximations c_i to c_0 on a dyadic resolution scale using a low-pass filter h according to :

$$c_i(k) = \sum_u h(u) c_{i-1}(k + 2^{i-1}u) = \sum_u \frac{1}{2^i} \phi\left(\frac{k-u}{2^i}\right) c_0(u) \quad (12)$$

where $h = \{1/16, 1/4, 3/8, 1/4, 1/16\}$ is actually the set of coefficients in the scaling equation for the cubic spline :

$$\phi(k) = \sum_u h(u) \phi(2k - u) \quad (13)$$

We note that each c_i is the same size as the original data c_0 and that the lowest resolution J_{max} is obviously limited by data size T . Then, taking the difference between two consecutive approximations gives the details at that scale or the wavelet coefficients

$$w_i(k) = c_{i-1}(k) - c_i(k) = \sum_u \frac{1}{2^{i-1}} \psi\left(\frac{k-u}{2^{i-1}}\right) c_0(u) \quad (14)$$

where the wavelet function $\psi(k)$ is defined by :

$$\psi(k) = \phi(k) - \frac{1}{2} \phi\left(\frac{k}{2}\right) \quad (15)$$

The w_i 's and c_i 's given using the *à trous* algorithm actually are obtained by passing the original signal c_0 through a set of finite impulse response (FIR) filters $\psi_1, \psi_2, \dots, \psi_J, \phi_J$. An essential property of these filters is that an inverse transform exists. In fact, reconstruction results simply from adding all the wavelet scales together with the last smooth approximation :

$$\forall k, c_0(k) = c_J(k) + w_J(k) + w_{J-1}(k) + \dots + w_2(k) + w_1(k) \quad (16)$$

The above *à trous* algorithm is easily extendable to two-dimensional images :

$$c_i(k, l) = \sum_u \sum_v h(u, v) c_{i-1}(k + 2^{i-1}u, l + 2^{i-1}v) \quad (17)$$

$$w_i(k, l) = c_{i-1}(k, l) - c_i(k, l) \quad (18)$$

and the reconstruction is still a simple co-addition of the wavelet scales and the smooth array :

$$c_0(k, l) = c_J(k, l) + \sum_{i=1}^J w_i(k, l) \quad (19)$$

The use of the B_3 spline leads to a convolution with the 5×5 mask h :

$$\frac{1}{256} \begin{pmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix}$$

but it is faster to compute the convolution in a separable way (first on rows, and then on the resulting columns).

3.2 Spectral matching in wavelet space : wSMICA

Consider the set of ideal band pass filters \mathcal{F}_q associated with non-overlapping frequency domains F_q as used by the Fourier space implementation of SMICA. Let Y_q denote the stationary Gaussian random processes obtained by passing the observations X of size m through filter \mathcal{F}_q . Let \tilde{Y}_q be their Fourier coefficients. Because of the unitary property of the Fourier transform, considering a batch of T samples $X_{t=1,T}$, the following equality between joint probabilities holds :

$$P(Y_{1;t=1,T}, \dots, Y_{Q;t=1,T}) = P(\tilde{Y}_{1;k=1,T}, \dots, \tilde{Y}_{Q;k=1,T}) \quad (20)$$

Assuming uncorrelated Fourier coefficients as in the above mentioned maximum likelihood derivation of SMICA based on the the Whittle approximation, and because of the non-overlapping filters, it follows that the $Y_{q;t}$ for different q 's are also decorrelated so that :

$$-\log P(Y_{1;t=1,T}, \dots, Y_{Q;t=1,T}) = -\sum_{q=1}^Q \log P(\tilde{Y}_{q;k=1,T}) \quad (21)$$

and that $\forall q$:

$$\begin{aligned} -\log P(Y_{q;k=1,T}) &= -\log P(\tilde{Y}_{q;k=1,T}) \\ &= n_q \mathcal{D}_{KL} \left(\hat{R}_{X,q}^f, AR_{S,q}^f A^\dagger + R_{N,q}^f \right) \end{aligned} \quad (22)$$

Now define mixture, source and noise covariances $R_{X,q}^t$, $R_{S,q}^t$ and $R_{N,q}^t$ in the time domain at the output of the above filters. The former matrices can be estimated from the available data using :

$$\hat{R}_{X,q}^t = \frac{1}{T} \sum_{t=0}^{T-1} Y_{q;t} Y_{q;t}^\dagger \quad (23)$$

and nothing opposes attempting component separation by spectral matching in the time domain using these latter covariances by minimizing

$$\phi(\theta) = \sum_{q=1}^Q \alpha_q \mathcal{D} \left(\hat{R}_{X,q}^t, AR_{S,q}^t A^\dagger + R_{N,q}^t \right) \quad (24)$$

with respect to $\theta = (A, R_{S,q}^t, R_{N,q}^t)$, provided the estimated covariances are full rank matrices. However, deriving adequate weights α_q in order to get a good approximation of the likelihood is not straightforward because of the correlations between the $Y_{q;t}$'s at different t 's. In fact, owing to these correlations, the convergence of $\hat{R}_{X,q}^t$ to $R_{X,q}^t$ can be very slow. The helpful point equation (22) actually makes is that taking $\alpha_q = n_q$ will correctly reflect our confidence in the estimated covariances $\hat{R}_{X,q}^t$.

The next step is obviously to use another set of filters in place of the ideal band pass filters used by SMICA. In fact, in dealing with non stationary data or, as a special case, with gapped data, it is especially attractive to consider finite impulse response filters. Indeed, provided the response of such a filter is short enough compared to data size T and gap widths, not all the samples in the filtered signal will be affected by the gaps. Therefore, using these latter samples exclusively, one may expect better estimation of the statistical properties of the original data *i.e.* without the gaps. We choose in what follows to use filters $\psi_1, \psi_2, \dots, \psi_J, \phi_J$ (see figure 1) and the wavelet *à trous* algorithm described previously. An immediate consequence of this choice is that the decorrelation between the different

filter outputs no longer holds, due to their overlapping responses in Fourier space. However, we do benefit from the fast filtering algorithms and, which is quite significant, from the possibility of reconstructing estimated source templates.

Let us consider again a batch of T regularly spaced data samples $X_{t=1 \rightarrow T}$. Possible gaps in the data are simply described with a mask μ *i.e.* a vector of zeroes and ones the same length as X with ones corresponding to samples outside the gaps. Denoting W_1, W_2, \dots, W_J and C_J the wavelet scales and smooth approximation of X , obtained with the *à trous* transform and μ_1, \dots, μ_{J+1} the masks for the different scales determined from the original mask $\mu(t)$ knowing the different filter lengths, wavelet covariances are estimated as follows :

$$\begin{aligned}\hat{R}_{X,1 \leq i \leq J}^w &= \frac{1}{l_i} \sum_{t=1}^T \mu_i(t) W_i(t) W_i(t)^\dagger \\ \hat{R}_{X,J+1}^w &= \frac{1}{l_{J+1}} \sum_{t=1}^T \mu_{J+1}(t) C_J(t) C_J(t)^\dagger\end{aligned}\quad (25)$$

where l_i is the number of non zero samples in μ_i . With source and noise covariances $R_{S,i}^w, R_{N,i}^w$ defined in a similar way, the covariance model in wavelet space becomes

$$R_{X,i}^w = A R_{S,i}^w A^\dagger + R_{N,i}^w \quad (26)$$

and minimizing

$$\phi(\theta) = \sum_{q=1}^Q \alpha_q \mathcal{D}(\hat{R}_{X,q}^w, A R_{S,q}^w A^\dagger + R_{N,q}^w) \quad (27)$$

with respect to the model parameters $\theta_w = (A, R_{S,i}^w, R_{N,i}^w)$ achieves the desired component separation.

However, in order for $\phi(\theta)$ to be a good approximation to the likelihood, the weights α_q again have to be determined with care. These weights should account for the correlations between wavelet coefficients from different or the same scales, especially in the lower frequencies. Actually, exaggerating the so-called decorrelating property of the wavelet transform, we assume coefficients from different scales are uncorrelated. Nevertheless, coefficients from one same scale are strongly correlated, especially with the adopted *à trous* redundant transform. Then, in the case of complete data sets *i.e.* without gaps, and because the 1D wavelet filter length in the time domain doubles from scale to scale, the transposition of equation (22) leads to taking :

$$\{\alpha_1, \alpha_2, \dots, \alpha_J, \alpha_{J+1}\} = \left\{ \frac{1}{2}, \frac{1}{4}, \dots, \frac{1}{2J}, \frac{1}{2J} \right\} \quad (28)$$

In the 2D case, this becomes :

$$\{\alpha_1, \alpha_2, \dots, \alpha_J, \alpha_{J+1}\} = \left\{ \frac{3}{4}, \frac{3}{16}, \dots, \frac{3}{4J}, \frac{1}{4J} \right\} \quad (29)$$

However, when there are gaps in the data, the Fourier modes can be strongly correlated and the Whittle approximation is no longer appropriate. In order to derive an approximate likelihood function, consider the orthogonal discrete wavelet transform. In the 1D case, this is a non-redundant transform in which the number of coefficients is halved from scale to scale. It is common and quite convenient to assume these coefficients are uncorrelated. Denoting l_i^{DWT} the number of DWT coefficients unaffected by the gaps in scale i , these have the same statistical significance or information content

as the $l_i \approx 2^i \times l_i^{DWT}$ coefficients in scale i determined with the *à trous* wavelet transform. Finally, a good approximation to the likelihood is obtained taking

$$\{\alpha_1, \alpha_2, \dots, \alpha_J, \alpha_{J+1}\} = \left\{ \frac{l_1}{2}, \frac{l_2}{4}, \dots, \frac{l_J}{2J}, \frac{l_{J+1}}{2J} \right\} \quad (30)$$

or, in the 2D case, :

$$\{\alpha_1, \alpha_2, \dots, \alpha_J, \alpha_{J+1}\} = \left\{ \frac{3l_1}{4}, \frac{3l_2}{16}, \dots, \frac{3l_J}{4J}, \frac{l_{J+1}}{4J} \right\} \quad (31)$$

in equation (27). We will refer to this combination of principles from SMICA and wavelet transforms as wSMICA.

A point to be stressed here is that the number of bands in the case of wSMICA is very much limited by the original data size, which is not as strongly the case with SMICA. But this limitation is mostly a requirement for reconstruction using (11) and (16) to make sense. If the mixing matrix A is a parameter of greater interest and if there is no real need to estimate source maps S , then there is no objection in principle to using more redundant transforms such as the continuous wavelet transform, or in fact any set of linear filters (of finite impulse response to cope easily with edges and gaps). This in turn raises the question of optimally choosing this set of filters as in [12].

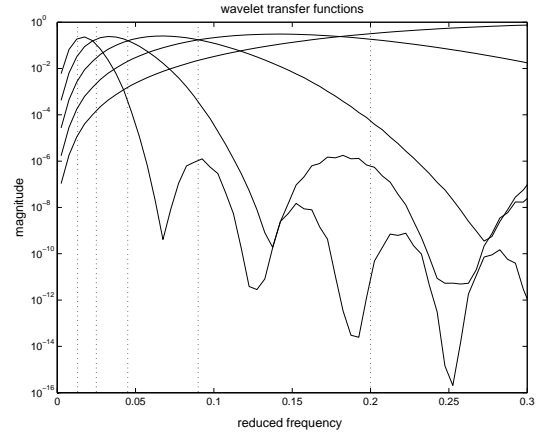


FIG. 1 – Magnitudes of the cubic spline wavelet filters $\psi_1, \psi_2, \dots, \psi_5$ used in the simulations described further down. The vertical dotted lines for $\nu = \{0.013, 0.025, 0.045, 0.09, 0.2, 0.5\}$ delimit the five frequency bands used with SMICA in these simulations.

4. NUMERICAL EXPERIMENTS

4.1 Simulated data

The methods described above were applied to synthetic observations consisting of $m = 6$ mixtures of $n = 3$ components namely CMB, galactic dust and SZ emissions for which typical templates, shown on figure 2, were obtained as described in [1].

The templates, and thus the mixtures in each simulated data set, consist of 300×300 pixel maps corresponding to a $12.5^\circ \times 12.5^\circ$ field located at high galactic latitude. The six mixtures in each set mimic observations that will eventually be acquired in the six frequency channels of the Planck-HFI on part-sky, local maps. The entries of the mixing matrix A used in these simulations actually are estimated values of the electromagnetic emission laws of the

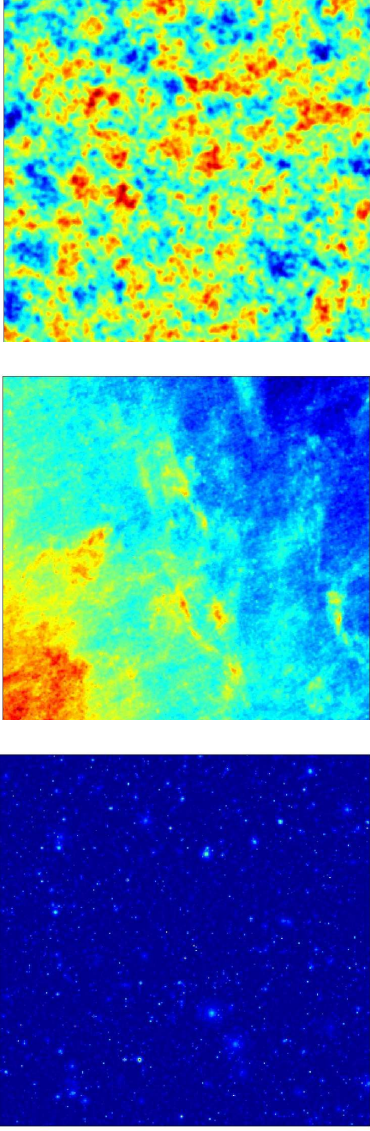


FIG. 2 – Simulated component templates for CMB (*top*), DUST (*middle*), SZ (*bottom*).

original components at 100, 143, 217, 353, 545 and 857 GHz. These values are grouped in table 1.

CMB	DUST	SZ	channel
7.452×10^{-1}	3.654×10^{-2}	-8.733×10^{-1}	100 GHz
5.799×10^{-1}	7.021×10^{-2}	-4.689×10^{-1}	143 GHz
3.206×10^{-1}	1.449×10^{-1}	-2.093×10^{-3}	217 GHz
7.435×10^{-2}	3.106×10^{-1}	1.294×10^{-1}	353 GHz
6.009×10^{-3}	5.398×10^{-1}	2.613×10^{-2}	545 GHz
6.115×10^{-5}	7.648×10^{-1}	5.268×10^{-4}	857 GHz

TAB. 1 – Entries of A , the mixing matrix used in our simulations.

White Gaussian noise was added to the mixtures according to equation (2) in order to simulate instrumental noise. While the relative noise standard deviations between channels were set according to the nominal values of the Planck HFI, we experimented five *global* noise levels at -20 ,

-6 , -3 , 0 and $+3$ dB from nominal values. Table 2 gives the typical energy fractions that are contributed by each of the $n = 3$ original sources and noise, to the total energy of each of the $m = 6$ mixtures, considering Planck nominal noise variance. In fact, because SMICA and wSMICA actually work on spectral bands, a much better indication of signal to noise ratio in these simulations is given by figure 3 where it is shown how noise and source energy contributions distribute with respect to frequency in the six mixtures.

CMB	DUST	SZ	noise	channel
9.91×10^{-1}	1.18×10^{-4}	7.92×10^{-3}	2.53×10^{-6}	100 GHz
9.97×10^{-1}	7.25×10^{-4}	3.79×10^{-3}	5.17×10^{-7}	143 GHz
9.98×10^{-1}	1.01×10^{-2}	2.48×10^{-7}	1.34×10^{-7}	217 GHz
5.55×10^{-1}	4.8×10^{-1}	9.78×10^{-3}	7.47×10^{-8}	353 GHz
2.5×10^{-3}	1.0	2.75×10^{-4}	3.78×10^{-9}	545 GHz
1.29×10^{-7}	1.0	5.56×10^{-8}	1.24×10^{-10}	857 GHz

TAB. 2 – Energy fraction contributed by each source to the total energy of each mixture, for the nominal noise variance on the Planck HFI channels.

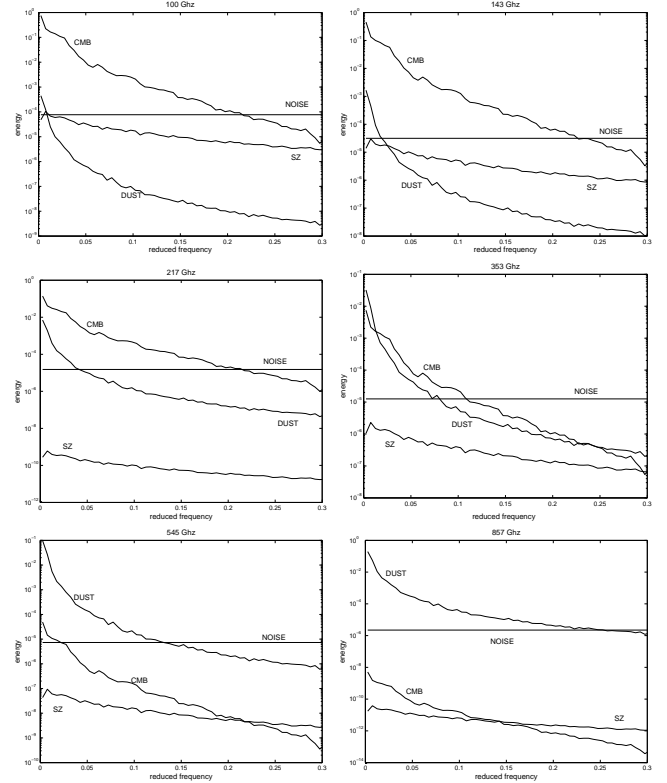


FIG. 3 – Energy contributed by each source and noise to each of the six mixtures (*mixture 1* : *top left*, *mixture 6* : *bottom right*) as a function of frequency, for the nominal noise variance on the Planck HFI channels. Note how SZ is expected to always be below nominal noise, that CMB and dust strongly dominate in different channels and that CMB and dust spectra, without being proportional, display the same general behaviour dominated by low modes.

Finally, in order to investigate the benefits of using wSMICA in place of SMICA when gaps are inserted in the data, the mask shown on figure 4 was applied onto the mixture maps. The case where no data is missing was also considered for the sake of comparison. In each of these two

particular configurations, spectral matching was assessed and optimized both at the output of the five wavelet filters ψ_1, \dots, ψ_5 associated to higher frequency details, and on the corresponding five bands in Fourier space, as shown on figure 1. This latter choice of frequency bands is simply made to ease comparison between SMICA and wSMICA. It may be argued that this choice is probably not optimal to run SMICA. But, in fact, the optimal selection of filters is clearly a meaningful question both for SMICA and wSMICA. This will require further investigation.

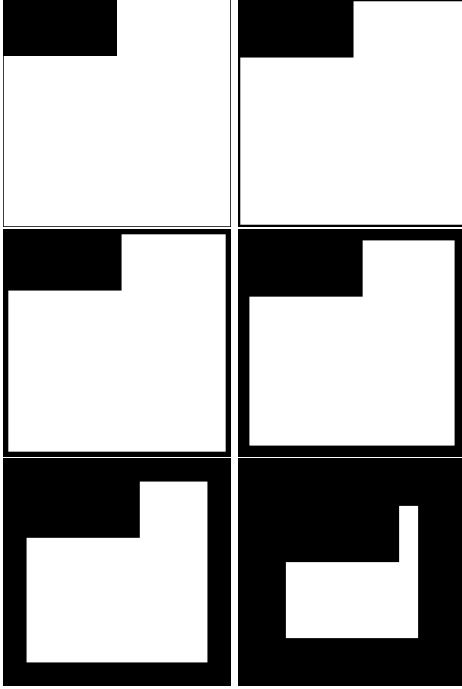


FIG. 4 – Mask used to simulate a gap in the data (*top left*), and the modified masks at scales 1 (*top right*) through 5 (*bottom left*). The discarded pixels are in black.

4.2 Preliminary results

Preliminary experiments were conducted in the case of vanishing instrumental noise variance, with a square 3×3 mixing matrix. It was mentioned before that in this limit, the spectral matching objective boils down to the joint diagonalization of covariance matrices. Further, taking the mixing matrix to be the identity matrix (*i.e.* try to separate sources which are not actually mixed), it is possible to gain some insight on the spectral diversity of the independent sources, for a given choice of bands or filters. Indeed, the performance of the independent component separation methods based on spectral matching depend highly on spectral diversity.

The following steps were repeated 1000 times :

- randomly pick one of each component maps out of the available 200 CMB maps, 30 dust maps and 1500 SZ maps.
- calculate covariance matrices in the five wavelet or Fourier bands, both with and without masking part of the maps, as is all described above.
- normalize each source so that its total energy over the five bands is equal to one.

- use the algorithm in [8] to jointly diagonalize the covariances in each configuration, and keep the resulting separating matrices.

If the sources have satisfactory spectral properties, the obtained separating matrices should not depart drastically from the identity matrix. Moreover, denoting \mathcal{A} any invertible 3×3 mixing matrix, and $\hat{\mathcal{A}}^{-1}$ the resulting separating matrix, it is shown in [14] that the variances of the off-diagonal terms in $\hat{\mathcal{A}}^{-1}\mathcal{A}$ depend only on spectral diversity, in the case of Gaussian sources. In fact, to assess the effect of any non-Gaussianity or non-stationarity in the source templates, the same experiment was repeated on Gaussian maps generated with the same spectra as the CMB, Dust and SZ components. In any case, the independent source components are separated using :

$$\hat{S} = \hat{\mathcal{A}}^{-1}\mathcal{A}S = \mathcal{I}S \quad (32)$$

so that with the above normalization, the square of any off-diagonal term \mathcal{I}_{ij} is directly related to the residual level of component j in the recovered component i .

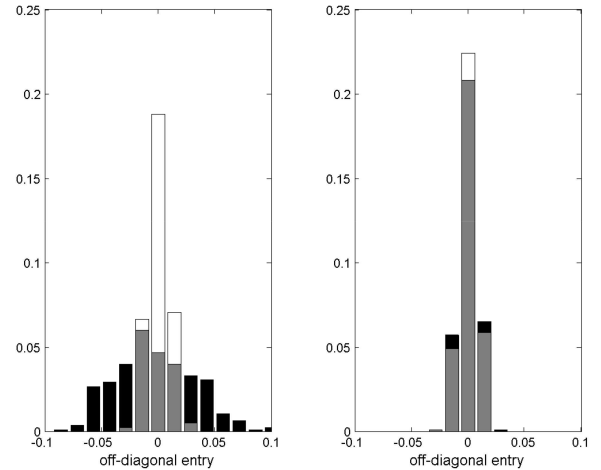


FIG. 5 – Histograms of the off diagonal term corresponding to the residual corruption of "CMB" by "Dust" while separating Gaussian maps generated with the same power spectra as the astrophysical components, by joint diagonalization of covariance matrices in Fourier (*left*) and wavelet (*right*) space, with (*black, which appears grey when seen through white*) and without (*white*) masking part of the the data. The dark widest histogram on the left highlights the impact of masking on source separation based on Fourier covariances.

The histograms on figure 5 are for the off diagonal term corresponding to the residual corruption of CMB by *Gaussian* Dust in the second set of experiments. In tables 3 and 4, the results obtained with the synthetic component maps are given as well as those obtained with the Gaussian maps, in terms of the standard deviations of the off-diagonal entries \mathcal{I}_{ij} defined by (32).

Interestingly, when working on Gaussian maps without masks, using covariances in Fourier space or in wavelet space gives similar performances. It is also satisfactory, when covariances in wavelet space are used with Gaussian maps, that each computed standard deviation only slightly increases when a mask is applied on the data. Indeed, as a consequence of incomplete coverage, there are less samples

	NM	M	Han
$\mathcal{I}_{1,2}$	0.097 <u>0.0076</u>	0.074 <u>0.038</u>	0.024
$\mathcal{I}_{1,3}$	0.0049 <u>0.0044</u>	0.005 <u>0.006</u>	0.0094
$\mathcal{I}_{2,1}$	0.017 <u>0.0066</u>	0.018 <u>0.01</u>	0.017
$\mathcal{I}_{2,3}$	0.0064 <u>0.0077</u>	0.0066 <u>0.0096</u>	0.011
$\mathcal{I}_{3,1}$	0.0024 <u>0.0026</u>	0.0028 <u>0.0037</u>	0.0039
$\mathcal{I}_{3,2}$	0.0054 <u>0.0071</u>	0.0054 <u>0.0079</u>	0.01

TAB. 3 – Standard deviations of the off-diagonal entries \mathcal{I}_{ij} defined by (32) obtained while separating realistic component maps by joint diagonalization of covariance matrices in **Fourier** space, with (M) or without masking (NM) part of the data, or applying an apodizing Hanning window (Han). Components 1, 2 and 3 respectively stand for CMB, Dust and SZ. The numbers in *italic* were obtained with Gaussian maps and the underlined numbers correspond to the histograms in figure 5.

	NM	M
$\mathcal{I}_{1,2}$	0.015 <u>0.0071</u>	0.018 <u>0.0079</u>
$\mathcal{I}_{1,3}$	0.0025 <u>0.0029</u>	0.0028 <u>0.0031</u>
$\mathcal{I}_{2,1}$	0.016 <u>0.0077</u>	0.019 <u>0.0089</u>
$\mathcal{I}_{2,3}$	0.0041 <u>0.0051</u>	0.0048 <u>0.0075</u>
$\mathcal{I}_{3,1}$	0.0024 <u>0.0029</u>	0.003 <u>0.0039</u>
$\mathcal{I}_{3,2}$	0.0039 <u>0.0054</u>	0.0053 <u>0.0085</u>

TAB. 4 – Standard deviations of the off-diagonal entries \mathcal{I}_{ij} defined by (32) obtained while separating realistic component maps by joint diagonalization of covariance matrices in **wavelet** space, with (M) and without masking (NM) part of the data. Components 1, 2 and 3 respectively stand for CMB, Dust and SZ. The numbers in *italic* were obtained with Gaussian maps and the underlined numbers correspond to the histograms in figure 5.

from which to estimate the covariances. This increase is also observed when covariances in Fourier space are used with the Gaussian maps but it can be as high as five-fold and it does not affect all coefficients the same way. Although this can again be attributed to the reduced data size, the

lowered spectral diversity between components, because of the correlations and smoothing induced in Fourier space by the mask, is also part of the explanation. In fact, as shown on figure 3, CMB and dust spatial power spectra are somewhat similar, *i.e.* show low spectral diversity, and further smoothing can only degrade the performance of the source separation algorithm based on Fourier covariances.

In the case of realistic component maps, we note first that the comparison of the performance of component separation using wavelet covariances with and without mask again agrees with the different data sizes, which is not the case with covariances in Fourier space. Next, whether covariances in Fourier or wavelet space are used, we note that the terms coupling CMB and Dust are again much higher in magnitude, even on complete maps. It seems that the actual non-stationarity and non-Gaussianity of the realistic component maps are relevant issues. Another point is that the CMB and Dust templates as in figure 2 exhibit sharp edges compared to SZ and this inevitably disturbs spectral estimation using a simple DFT. To assess this effect, simulations were also conducted where the covariances in Fourier space were computed after an apodizing Hanning window was applied on the complete data maps. The results reported in table 3, to be compared to table 4, do indicate a slightly positive effect of windowing, but still the separation using wavelet covariances appears better.

4.3 Realistic experiments

The above preliminary results clearly point out in the noiseless case the advantageous use of wavelets to easily escape the very bad impact that gaps and sharp edges actually have on the performance of the source separation using covariances in Fourier space. Hence this is strong encouragement to move on to investigating the effect of additive noise on the mixture maps according to (2), using SMICA and its extension wSMICA. We note that although in the case of wSMICA the link with maximum likelihood is not as strongly asserted as with SMICA, the optimization algorithm used in the simulations hereafter consists in both cases of the same heuristic succession of EM and BFGS steps and initialization is done as discussed in paragraph 2.2.

Picking at random one of each component maps out of the available 200 CMB maps, 30 dust maps and 1500 SZ maps, 1000 synthetic mixture maps were generated as previously described, for each of the 5 noise levels chosen. Then, component separation was conducted using the spectral matching algorithms SMICA and wSMICA both with and without part of the maps being masked. Now, each run of SMICA and wSMICA on the data returns estimates \hat{A}_f and \hat{A}_w of the mixing matrix. Clearly, these estimates are subject to the indeterminacies inherent to the instantaneous linear mixture model (2). Indeed, in the case where optimization is over all parameters θ , it is obvious that any simultaneous permutation of the columns of A and of the lines of S leaves the model unchanged. The same occurs when exchanging a scalar possibly negative factor between any column in A and the corresponding line in S . Therefore, columnwise comparison of \hat{A}_f and \hat{A}_w to the original mixing matrix A requires first fixing these indeterminacies. This is done by hand after \hat{A}_f and \hat{A}_w have been normalized columnwise.

The results we report next concentrate on the statistical properties of \hat{A}_f and \hat{A}_w as estimated from the 1000 runs of the two competing methods in the several configurations

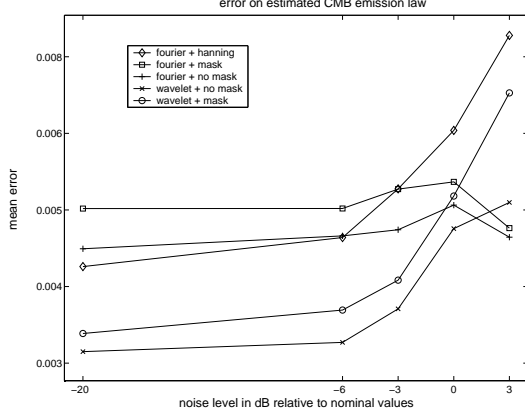


FIG. 6 – Comparison of the mean squared errors on the estimation of the emissivity of **CMB** as a function of noise in five different configurations namely : wSMICA without mask, wSMICA with mask, fSMICA without mask, fSMICA with mask, fSMICA with Hanning apodizing window.

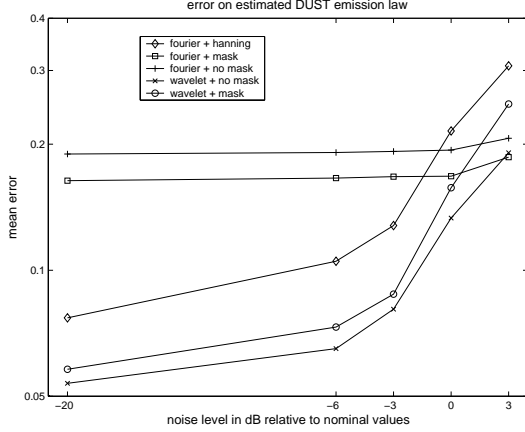


FIG. 7 – Comparison of the mean squared errors on the estimation of the emissivity of **DUST** as a function of noise in five different configurations namely : wSMICA without mask, wSMICA with mask, fSMICA without mask, fSMICA with mask, fSMICA with Hanning apodizing window.

retained. In fact, the correct estimation of the mixing matrix in model (2) is a relevant issue for instance when it comes to dealing with the cross calibration of the different detectors. Figures 6, 7 and 8 show the results obtained, using the quadratic norm

$$QE_j = \left(\sum_{i=1}^m (A_{ij} - \hat{A}_{ij})^2 \right)^{\frac{1}{2}} \quad (33)$$

with $\hat{A} = \hat{A}_f$ or \hat{A}_w and $j = \text{CMB, DUST or SZ}$, to assess the residual errors on the estimated emissivities of each component. The plotted curves show how the mean of the above positive error measure varies with increasing noise variance. For the particular case of CMB, table 5 gives the estimated standard deviations of the relative errors

$$\frac{A_{ij} - \hat{A}_{ij}}{A_{ij}} \quad (34)$$

on the estimated CMB emissivity in the six channels of

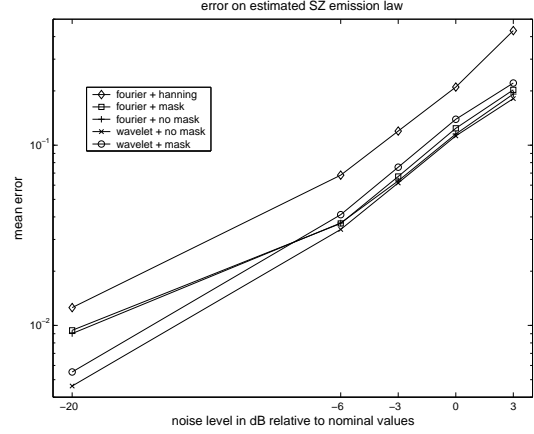


FIG. 8 – Comparison of the mean squared errors on the estimation of the emissivity of **SZ** as a function of noise in five different configurations namely : wSMICA without mask, wSMICA with mask, fSMICA without mask, fSMICA with mask, fSMICA with Hanning apodizing window.

Planck's HFI in the different configurations retained.

	WNM	WM	FNM	FM	FHan
\mathcal{A}_{11}	4.4×10^{-4}	5.0×10^{-4}	6.2×10^{-4}	7.3×10^{-4}	7.2×10^{-4}
	5.4×10^{-4}	7.5×10^{-4}	7.1×10^{-4}	8.5×10^{-4}	9.5×10^{-4}
	6.6×10^{-4}	9.2×10^{-4}	8.2×10^{-4}	8.9×10^{-4}	1.3×10^{-3}
	9.4×10^{-4}	1.2×10^{-3}	1.0×10^{-3}	1.0×10^{-3}	1.7×10^{-3}
	1.2×10^{-3}	1.7×10^{-3}	1.2×10^{-3}	1.4×10^{-3}	2.3×10^{-3}
\mathcal{A}_{21}	1.6×10^{-4}	2.1×10^{-4}	2.1×10^{-4}	2.0×10^{-4}	2.7×10^{-4}
	5.3×10^{-4}	7.8×10^{-4}	5.6×10^{-4}	5.7×10^{-4}	1.0×10^{-3}
	7.0×10^{-4}	1.1×10^{-3}	7.6×10^{-4}	8.4×10^{-4}	1.4×10^{-3}
	1.0×10^{-3}	1.6×10^{-3}	1.0×10^{-3}	1.0×10^{-3}	2.1×10^{-3}
	1.4×10^{-3}	2.2×10^{-3}	1.5×10^{-3}	1.7×10^{-3}	3.1×10^{-3}
\mathcal{A}_{31}	1.5×10^{-3}	1.8×10^{-3}	2.2×10^{-3}	2.5×10^{-3}	2.3×10^{-3}
	1.7×10^{-3}	2.1×10^{-3}	2.3×10^{-3}	2.6×10^{-3}	2.9×10^{-3}
	2.1×10^{-3}	2.6×10^{-3}	2.6×10^{-3}	2.8×10^{-3}	3.7×10^{-3}
	2.7×10^{-3}	3.0×10^{-3}	2.9×10^{-3}	3.0×10^{-3}	4.2×10^{-3}
	3.3×10^{-3}	4.6×10^{-3}	3.3×10^{-3}	3.5×10^{-3}	6.1×10^{-3}
\mathcal{A}_{41}	1.8×10^{-2}	2.0×10^{-2}	2.7×10^{-2}	3.0×10^{-2}	2.5×10^{-2}
	1.9×10^{-2}	2.1×10^{-2}	2.7×10^{-2}	2.1×10^{-2}	2.7×10^{-2}
	2.1×10^{-2}	2.4×10^{-2}	2.8×10^{-2}	3.1×10^{-2}	2.9×10^{-2}
	2.7×10^{-2}	2.8×10^{-2}	3.1×10^{-2}	3.0×10^{-2}	3.5×10^{-2}
	3.0×10^{-2}	4.1×10^{-2}	2.5×10^{-2}	2.7×10^{-2}	4.9×10^{-2}
\mathcal{A}_{51}	4.0×10^{-1}	4.5×10^{-1}	6.1×10^{-1}	6.6×10^{-1}	5.6×10^{-1}
	4.2×10^{-1}	4.7×10^{-1}	6.1×10^{-1}	6.5×10^{-1}	5.8×10^{-1}
	4.5×10^{-1}	5.0×10^{-1}	6.1×10^{-1}	6.7×10^{-1}	6.4×10^{-1}
	5.7×10^{-1}	5.9×10^{-1}	6.7×10^{-1}	6.7×10^{-1}	7.5×10^{-1}
	6.2×10^{-1}	8.4×10^{-1}	5.0×10^{-1}	5.5×10^{-1}	1.0
\mathcal{A}_{61}	5.7×10^1	6.2×10^1	8.5×10^1	9.2×10^1	7.8×10^1
	5.8×10^1	6.5×10^1	8.6×10^1	9.1×10^1	8.1×10^1
	6.2×10^1	6.9×10^1	8.6×10^1	9.4×10^1	8.9×10^1
	7.9×10^1	8.2×10^1	9.3×10^1	9.2×10^1	1.0×10^2
	8.6×10^1	1.2×10^2	6.9×10^1	7.7×10^1	1.4×10^2

TAB. 5 – **Standard deviations** of the relative errors on the estimated emissivities \mathcal{A}_{i1} of CMB in Planck's HFI six channels. The column labels WNM, WM, FNM, FM, FHan are for the different configurations, respectively : wSMICA without mask, wSMICA with mask, fSMICA without mask, fSMICA with mask, fSMICA with Hanning apodizing window. The five figures in each box are for noise variance -20, -6, -3, 0 and 3 dB from nominal Planck values.

Closer to our source separation objective, a more significant way of assessing the quality of \hat{A}_f and \hat{A}_w as estimators of the mixing matrix A , would be to use the following signal to interference ratio :

$$ISR_j = \frac{\mathcal{I}_{j,j}^2 \sigma_j^2}{\sum_{i \neq j} \mathcal{I}_{j,i}^2 \sigma_i^2} \quad (35)$$

where the σ_j are the source variances and

$$\mathcal{I} = (\hat{A}^\dagger \hat{R}_N^{-1} \hat{A})^{-1} \hat{A}^\dagger \hat{R}_N^{-1} A \quad (36)$$

with R_N the noise covariance. The plots on figures 9, 10 and 11 show how the mean ISR from the 1000 runs of SMICA and wSMICA in different configurations, varies with increasing noise.

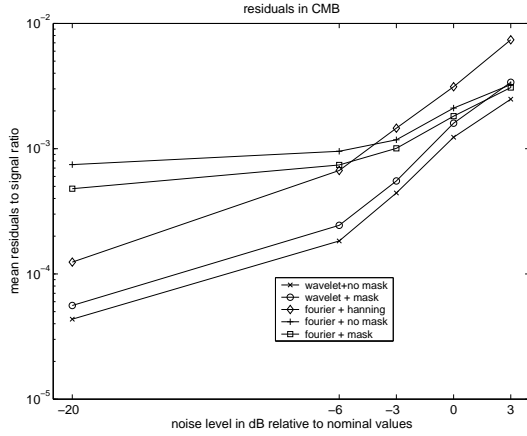


FIG. 9 – Comparison of the mean ISR for **CMB** as a function of noise in five different configurations namely : wSMICA without mask, wSMICA with mask, fSMICA without mask, fSMICA with mask, fSMICA with Hanning apodizing window.

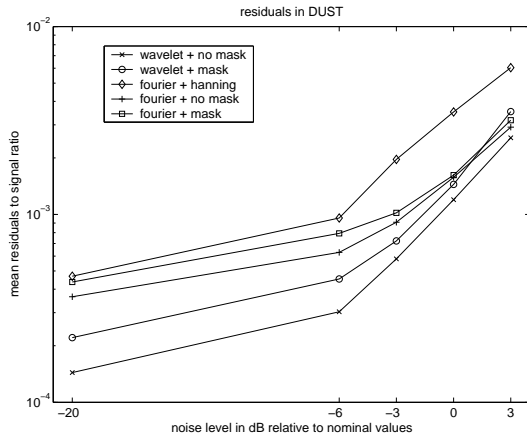


FIG. 10 – Comparison of the mean ISR for **DUST** as a function of noise in five different configurations namely : wSMICA without mask, wSMICA with mask, fSMICA without mask, fSMICA with mask, fSMICA with Hanning apodizing window.

We note again that the performance of wSMICA behaves as expected when noise increases and if part of the data is

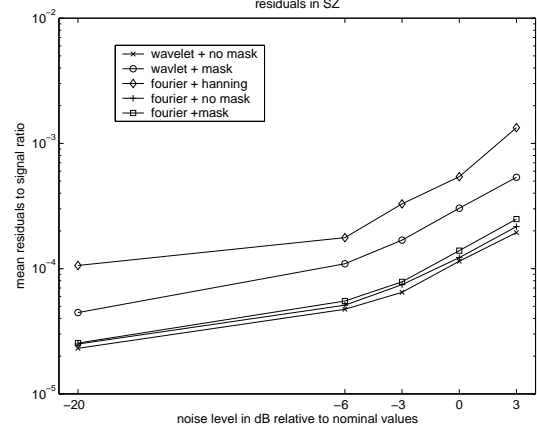


FIG. 11 – Comparison of the mean ISR for **SZ** as a function of noise in five different configurations namely : wSMICA without mask, wSMICA with mask, fSMICA without mask, fSMICA with mask, fSMICA with Hanning apodizing window.

missing. However this is not always the case with SMICA. Finally this set of simulations, conducted in a more realistic setting with respect to ESA's Planck mission, again confirms the higher performance, over Fourier analysis, that we indeed expected from the use of wavelets. The latter are able to correctly grab the spectral content of partly masked data maps and from there allow for better component separation.

5. CONCLUSION

This paper has presented an extension of the Spectral Matching ICA algorithm to the case where the collected data is both correlated and non stationary, considering maps with gaps as a particular instance of practical significance. It was shown that simply substituting covariance matching in Fourier space by covariance matching in wavelet space enables to cope in the most general and straightforward way with gaps of possibly any shape. Mainly, it is the FIR nature of the wavelet filters used that allows the impact of edges and gaps on the estimated covariances and hence on component separation to be lowered. Optimally choosing the FIR filter-bank regarding a particular application is a possible further enhancement.

Results obtained with simulated astrophysical data as expected from the Planck mission were given and these confirm the benefits of correctly processing existing gaps. Clearly, other possible types of non-stationarities in the collected data such as spatially varying noise or component variance, etc. can be dealt with very simply in a similar fashion using the wavelet extension of SMICA.

In the CMB application, the mixed components have quite different statistical properties : some are expected to be very close to Gaussian whereas others are strongly non Gaussian. Standard ICA methods exploit the non Gaussianity of the mixed components. However, it is not clear yet how best to combine non Gaussianity and spectral diversity in order to perform better source separation. Other features of wavelets which are known to be powerful tools for the analysis and sparse representation of structured data might reveal useful here.

A. APPENDIX : EM ALGORITHM WITH CONSTRAINTS ON THE MIXING MATRIX

Considering Q separate frequency bands of size n_q with $\sum n_q = 1$, the EM functional derived for the instantaneous mixing model (2) with independent Gaussian stationary sources S and noise N is :

$$\Phi(\underline{\theta}, \theta) = \mathcal{E} \{ \log p(X, S | \underline{\theta}) | \theta \} \quad (37)$$

with $\theta = (A, R_{S,1}, \dots, R_{S,Q}, R_{N,1}, \dots, R_{N,Q})$ and $\underline{\theta} = (\underline{A}, \underline{R}_{S,1}, \dots, \underline{R}_{S,Q}, \underline{R}_{N,1}, \dots, \underline{R}_{N,Q})$. The maximization step of the EM algorithm seeks then to maximize $\Phi(\underline{\theta}, \theta)$ with respect to $\underline{\theta}$ and the optimal $\underline{\theta}$ is used as the value for θ at the next EM step, and so on until satisfactory convergence is reached. Explicit expressions are easily derived for the optimal $\underline{\theta}$ in the white noise case where an interesting decoupling occurs between the re-estimating equations for noise variances, source variances and the mixing matrix [11].

Linear equality constraints

When A is subject to linear constraints, the joint maximization of the EM functional with respect to all model parameters is no longer easily achieved in general. In fact, one cannot simply decouple the re-estimating rules for the noise parameters and the mixing matrix and these have to be optimized separately. We give next the modified re-estimating equations for the mixing matrix and the source variances in the case of constant noise (*i.e.* $\theta = (A, R_{S,1}, \dots, R_{S,Q})$).

First, let us exhibit the quadratic dependence of the EM functional $\Phi(\underline{\theta}, \theta)$ on \underline{A} :

$$\begin{aligned} \Phi(\underline{\theta}, \theta) = & -\frac{1}{2} \sum_q n_q \text{tr} \left(\underline{A} R_q^{ss} \underline{A}^\dagger R_{N,q}^{-1} \right. \\ & \left. - \underline{A} R_q^{xs\dagger} R_{N,q}^{-1} - R_q^{xs} \underline{A}^\dagger R_{N,q}^{-1} \right) + \text{const}_{\underline{A}} \end{aligned} \quad (38)$$

where

$$C_q = (A^\dagger R_{N,q}^{-1} A + R_{S,q}^{-1})^{-1} \quad (39)$$

$$W_q = (A^\dagger R_{N,q}^{-1} A + R_{S,q}^{-1})^{-1} A^\dagger R_{N,q}^{-1} \quad (40)$$

$$R_q^{xs} = \hat{R}_{X,q} W_q^\dagger \quad (41)$$

$$R_q^{ss} = W_q \hat{R}_{X,q} W_q^\dagger + C_q \quad (42)$$

In the white noise case, $R_{N,q} = R_N$, equation (38) becomes :

$$\begin{aligned} \Phi(\underline{\theta}, \theta) = & -\frac{1}{2} \text{tr} \left((\underline{A} - R^{xs} R^{ss-1}) R^{ss} \right. \\ & \left. (\underline{A} - R^{xs} R^{ss-1})^\dagger R_N^{-1} \right) + \text{const}_{\underline{A}} \end{aligned} \quad (43)$$

where :

$$R^{xs} = \sum_q n_q R_q^{xs} \quad \text{and} \quad R^{ss} = \sum_q n_q R_q^{ss} \quad (44)$$

Again, this can be re-written as :

$$\Phi(\underline{\theta}, \theta) = -\frac{1}{2} (\underline{A} - \mathcal{M}) \mathcal{Q} (\underline{A} - \mathcal{M})^\dagger + \text{const}_{\underline{A}} \quad (45)$$

where :

$$\underline{A} = \text{vect} \underline{A} \quad , \quad \mathcal{Q} = \underline{R}_N^{-1} \otimes \sum_q n_q R_q^{ss} \quad (46)$$

$$\mathcal{M} = \text{vect} \left(\left(\sum_q n_q R_q^{ys} \right) \left(\sum_q n_q R_q^{ss} \right)^{-1} \right) \quad (47)$$

With “vect”, we build a column vector with the entries of a matrix taken along its lines. Now let us consider linear constraints on the mixing matrix, specified as follows :

$$\mathcal{C}^\dagger (\underline{A} - \mathcal{A}_0) = 0 \quad (48)$$

where \mathcal{C} is a matrix with as many columns as constraints, and the columns of \mathcal{C} are the same size as \underline{A} . The maximum of the EM functional with respect to $\underline{\theta}$ subject to the specified linear constraints is then reached for :

$$\underline{A} = \mathcal{M} - \mathcal{Q} \mathcal{C} \left(\mathcal{C}^\dagger \mathcal{Q} \mathcal{C} \right)^{-1} \mathcal{C}^\dagger (\mathcal{M} - \mathcal{A}_0) \quad (49)$$

and

$$\underline{R}_{S,q} = \text{diag}(R_q^{ss}) \quad (50)$$

where “diag” returns a matrix with the same diagonal entries as its input argument.

In the free noise case, things are quite similar except that the noise covariance matrices $R_{N,q}$ do not factorize out as nicely. The EM functional is again expressed as :

$$\Phi(\underline{\theta}, \theta) = -\frac{1}{2} (\underline{A} - \mathcal{M}) \mathcal{Q} (\underline{A} - \mathcal{M})^\dagger + \text{const}_{\underline{A}} \quad (51)$$

where in this case :

$$\mathcal{Q} = \sum_q n_q R_{N,q}^{-1} \otimes R_q^{ss} \quad (52)$$

and

$$\mathcal{M} = \mathcal{Q}^{-1} \text{vect} \left(\sum_q n_q R_{N,q}^{-1} R_q^{xs} \right) \quad (53)$$

Then, the maximum of the EM functional with respect to $\underline{\theta}$ subject to the specified linear constraints is again reached for :

$$\underline{A} = \mathcal{M} - \mathcal{Q} \mathcal{C} \left(\mathcal{C}^\dagger \mathcal{Q} \mathcal{C} \right)^{-1} \mathcal{C}^\dagger (\mathcal{M} - \mathcal{A}_0) \quad (54)$$

and

$$\underline{R}_{S,q} = \text{diag}(R_q^{ss}) \quad (55)$$

These expressions of the re-estimates of the mixing matrix can become algorithmically very simple when for instance the linear constraints to be dealt with affect separate lines of A , or even simpler when the constraints are such that the entries of A are affected separately.

Positivity constraints on the entries of A

Suppose a subset of entries of A are constrained to be positive. The maximization step of the EM algorithm on A alone, again has to be modified. We suggest dealing with such constraints in a combinatorial way rephrasing the problem in terms of equality constraints. If the unconstrained maximum of the EM functional is not in the specified domain, then one has to look for a maximum on the borders of that domain : on a hyperplane, on the intersection of two, or three, or more hyperplanes. One important point is that the maximum of the EM functional with respect to A subject to a set of equality constraints will necessarily be lower than the maximum of the same functional considering any subset of these equality constraints. Hence, not all combinations need be explored, and a Branch and Bound type algorithm is well suited [13]. A straightforward extension allows to deal with the case where a set of entries of the mixing matrix are constrained by upper and lower bounds.

REFERENCES

- [1] J. Delabrouille, J.-F. Cardoso and G. Patanchon, "Multi-detector multi-component spectral matching and application for CMB data analysis", *Mon. Not. R. Astron. Soc.*, vol. 346, no 4, pp. 1089-1102, 2003.
- [2] C. Baccigalupi, L. Bedini et al., "Neural networks and separation of cosmic microwave background and astrophysical signals in sky maps", *Mon. Not. R. Astron. Soc.*, vol. 318, no 3, pp.769-780, 2000.
- [3] E. Kuruoglu, L. Bedini et al., "Source separation in astrophysical maps using independent factor analysis", *Neural netw.*, vol. 16, no 3-4, pp. 479-491, 2003.
- [4] D. Maino, A. Farusi et al., "All-sky astrophysical component separation with Fast Independent Component Analysis (FASTICA)", *Mon. Not. R. Astron. Soc.*, vol. 334, no 1, pp. 53-68, 2002.
- [5] H. Eriksen, A. Banday et al., "Foreground removal by an internal linear combination method : limitations and implications", *preprint astro-ph0401276*.
- [6] M. Hobson, A. Jones et al., "Foreground separation methods for satellite observations of the microwave background", *Mon. Not. R. Astron. Soc.*, vol. 300, no 1 pp. 1-29, 1998.
- [7] G. Patanchon, J. Delabrouille and J.-F. Cardoso, "Source separation on astrophysical data sets from the wmap satellite", *to appear in Proceedings of ICA2004*.
- [8] D.-T. Pham and P. Garat, "Blind separation of mixtures of independent sources through a quasi maximum likelihood approach", *IEEE Trans. Signal Processing*, vol. 45, no. 7, pp. 1712-1725, 1997.
- [9] J.-F. Cardoso, "The three easy routes to independent component analysis; contrasts and geometry", *Proc. of ICA2001 Workshop*, San Diego, 2001.
- [10] J.-L. Starck, F. Murtagh and A. Bijaoui, *Image and Data Analysis : The multiscale approach*, Cambridge University Press, 1998.
- [11] J.-F. Cardoso, H. Snoussi et al., "Blind separation of noisy Gaussian stationary sources. Application to cosmic microwave background imaging." *Proc. EUSIPCO 2002*, pp. 561-564, Toulouse (France), 2002.
- [12] P. Stoica, E. Larsson and J. Li, "Adaptive filter-bank approach to restoration and spectral analysis of gapped data", *The Astronomical Journal*, vol. 102, pp 2163-2173, October 2000.
- [13] P. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection", *IEEE Transactions on Computers*, vol. 26, no 9, pp 917-922, 1977.
- [14] J.-F. Cardoso, "Separation of non-stationary sources. Achievable performance", *Proceedings of SSAP2000*, pp 359-363, 2000.
- [15] M. Tegmark, D.-J. Eisenstein et al., "Foregrounds and forecasts for the cosmic microwave background", *The Astrophysical Journal*, 530, 133 (2000).
- [16] A. Benoit et al., "Archeops : a high resolution, large sky coverage balloon experiment for mapping cosmic microwave background anisotropies", *Astroparticle Physics*, vol. 17, no 2, pp 101-124, 2002.
- [17] P. de Bernardis et al., "A flat Universe from high-resolution maps of the cosmic microwave background radiation", *Nature*, no 404, pp 955-959, 2000.
- [18] S. Hanany et al., "MAXIMA=1 : a measurement of the cosmic microwave background anisotropy on angular scales of 10 arcminutes to 5 degrees", *Astrophysical Journal Letters*, 545, L5, 2000.
- [19] C. L. Bennett et al., "First year Wilkinson microwave anisotropy probe (Wmap) observations : preliminary maps and basic results", *ApJ. Suppl.*, vol. 148, 1, 2003.
- [20] G. Jungman et al., "Cosmological parameter determination with microwave background maps", *Phys. Rev. D*, 54, pp 1332-1344, 1996.
- [21] F. Bouchet and R. Gispert, "Foregrounds and CMB experiments : Semi-analytical estimates of contamination", *New Astronomy*, vol. 4, pp 443-479, 1999.
- [22] M. Tegmark and G. Efstathiou, "A method for subtracting foregrounds from multi-frequency CMB sky maps", *Mon. Not. R. Astron. Soc.*, vol. 281, pp 1297-1314, 1996.
- [23] M. P. Hobson , A. W. Jones et al., "Foreground separation methods for satellite observations of the cosmic microwave background", *Mon. Not. R. Astron. Soc.*, vol. 300, pp 1-29, 1998.